



# Macro vs. Micro Methods in Non-Life Claims Reserving (an Econometric Perspective)

Arthur Charpentier, Mathieu Pigeon

## ► To cite this version:

Arthur Charpentier, Mathieu Pigeon. Macro vs. Micro Methods in Non-Life Claims Reserving (an Econometric Perspective). 2016. hal-01280033

**HAL Id: hal-01280033**

**<https://hal.science/hal-01280033>**

Preprint submitted on 29 Feb 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# MACRO VS. MICRO METHODS IN NON-LIFE CLAIMS RESERVING (AN ECONOMETRIC PERSPECTIVE)

ARTHUR CHARPENTIER & MATHIEU PIGEON

## 1. INTRODUCTION

**1.1. Macro and Micro Methods.** For more than a century, actuaries have been using run-off triangles to project future payments, in non-life insurance. In the 30's, [1] formalized this technique that originated the popular *chain ladder* estimate. In the 90's, [2] proved that the *chain ladder* estimate can be motivated by a simple stochastic model, and later on [3] provided a comprehensive overview on stochastic models that can be connected with the *chain ladder* method, included regression models, that could be seen as extension of the so-called "factor" methods used in the 70's.

But using the terminology of [4] and [5], those were macro-level models for reserving. In the 70's, [6] suggested to used some marked point process of claims to project future payments, and quantify the reserves. More recently, [7],[8], [9], [10] or [11] (among many others) investigated further some probabilistic micro-level models. These models handle claims related data on an individual basis, rather than aggregating by underwriting year and development period. As mentioned in [12], these methods have not (yet) found great popularity in practice, since they are more difficult to apply.

All macro-level models are based on aggregate data found in a run-off triangle, which is their strength, but also probably their weakness. They are easy to understand, and can be mentioned in financial communication, without disclosing too much information. From a computational perspective, those models can also be implemented in a single spreadsheet. But recently, [3] started questioning actuaries about possible use of detailed micro-level information. Those models can incorporate heterogeneity, structural changes, etc (see [13] for a discussion).

---

*Date:* February, 28th, 2016.

Arthur Charpentier, UQAM-Quantact & Université de Rennes 1, France.  
Mathieu Pigeon, UQAM-Quantact.

**1.2. Best estimates and Variability.** In the context of macro-level models, [3] mention that prediction errors can be large, because of the small number of *observations* used in run-off triangles. Quantifying uncertainty in claim reserving methods is not only important in actuarial practice and to assess accuracy of predictive models, it is also a regulatory issue.

[14] and [15] obtained, on real data analysis, lower variance on the total amount of reserves with "micro" models than with "macro" ones. A natural question is about the generality of such result. Should "micro" model generate less variability than standard "macro" ones? That is the question that initiated that paper.

**1.3. Agenda.** In section 2, we detail intuitive results we expect when aggregating data by clusters, moving from micro-level models to macro-level ones. More precisely, we explain why with a linear model and a Poisson regression, macro- and micro-level models are equivalent. We also discuss the case of the Poisson regression model with random intercept. In section 3, we study "micro" and "macro" models in the context of claims reserving, on real data, as well as simulated ones.

## 2. CLUSTERING IN GENERALIZED LINEAR MIXED MODELS

In the economic literature, several papers discuss the use of "micro" vs. "macro" data, for instance in the context of unemployment duration in [16] or in the context of inflation in [17]. In [16], it is mentioned that both models are interesting, since "micro" data can be used to capture heterogeneity while "macro" data can capture cycle and more structural patterns. In [17], it is demonstrated that both heterogeneity and aggregation might explain the persistence of inflation at the macroeconomic level.

In order to clarify notation, and make sure that objects are well defined, we use small letters for sample values, e.g.  $y_i$ , and capital letters for underlying random variables, e.g.  $Y_i$  in the sense that  $y_i$  is a realisation of random variable  $Y_i$ . Hence, in the case of the linear model (see Section 2.1), we usually assume that  $Y_i \sim \mathcal{N}(\mathbf{x}_i^\top \mathbf{b}, \sigma^2)$ , and then  $\hat{\mathbf{b}}$  is the estimated model, in the sense that  $\hat{\mathbf{b}} = (\mathbf{x}\mathbf{x}^\top)^{-1}\mathbf{x}\mathbf{y}$  while  $\hat{\mathbf{B}} = (\mathbf{x}\mathbf{x}^\top)^{-1}\mathbf{x}\mathbf{Y}$  (here covariates  $\mathbf{x}$  are given, and non stochastic). Since  $\hat{\mathbf{B}}$  is seen as a random variable, we can write  $\mathbb{E}[\hat{\mathbf{B}}] = \mathbf{b}$ .

With a Poisson regression,  $Y_i \sim \mathcal{P}(\lambda_i)$  with  $\lambda_i = \exp[\mathbf{x}_i^\top \mathbf{b}]$ . In that case,  $\text{Var}[Y_i] = \mathbb{E}[Y_i] = \lambda_i$ . The estimate parameter  $\hat{\mathbf{b}}$  is a function of the observations,  $(\mathbf{x}_i, \mathbf{y}_i)$ 's, while  $\hat{\mathbf{B}}$  is a function of the observations,  $(\mathbf{x}_i, \mathbf{Y}_i)$ 's. In the context of the Poisson regression, recall that  $\mathbb{E}[\hat{\mathbf{B}}] \rightarrow$

$\mathbf{b}$  as  $n$  goes to infinity. With a quasi-Poisson regression,  $Y_i$  does not have, per se, a proper distribution. Nevertheless, its moments are well defined, in the sense that  $\text{Var}[Y_i] = \varphi \mathbb{E}[Y_i] = \varphi \lambda_i$ . And for convenience, we will denote  $Y_i \sim q\mathcal{P}(\lambda_i)$ , with an abuse of notation.

In this section, we will derive some theoretical results regarding aggregation in econometric models.

**2.1. The multiple linear regression model.** Consider a (multiple) linear regression model,

$$\begin{aligned} y_{i,g} &= \mathbf{x}_g^\top \mathbf{a} + \varepsilon_{i,g}, \\ \mathbf{a} &= [a_1 \quad \dots \quad a_{k+1}]^\top \quad \mathbf{x}_g = [x_{g,1} \quad \dots \quad x_{g,k+1}]^\top, \end{aligned} \tag{1}$$

where observations belong to a cluster  $g$  and are indexed by  $i$  within a cluster  $g$ ,  $i = 1, \dots, n_g$ ,  $g = 1, \dots, m$ . Assume further assumptions of the classical linear regression model [18], i.e.,

- (LRM1) no multicollinearity in the data matrix;
- (LRM2) exogeneity of the independent variables  $\mathbb{E}[\varepsilon_{i,g} | \mathbf{x}_g] = 0$ ,  $i = 1, \dots, n_g$ ,  $g = 1, \dots, m$ ; and
- (LRM3) homoscedasticity and nonautocorrelation of error terms with  $\text{Var}[\varepsilon_{i,g}] = \sigma^2$ .

Stacking observations within a cluster yield the following model

$$\bar{y}_g = \mathbf{x}_g^\top \mathbf{b} + e_g, \tag{2}$$

where

$$\bar{y}_g = \frac{1}{n_g} \sum_i y_{i,g} \text{ and } \mathbf{b} = [b_1 \quad \dots \quad b_{k+1}]^\top$$

with similar assumptions except for  $\text{Var}[e_g] = \sigma^2/n_g$ . Those two models are equivalent, in the sense that the following proposition holds.

**Proposition 2.1.** *Model (1) on a micro level and model (2) on a macro level are equivalent, in the sense that*

- (i)  $\hat{\mathbf{a}}_{OLS} = \hat{\mathbf{b}}_{OLS}$  when weights  $n_g$  are used in model (2); and
- (ii)  $\sum_{i,g} \hat{y}_{i,g} = \sum_g \hat{y}_g$  where  $y_g = n_g \bar{y}_g$ .

*Proof.* (i) The ordinary least-squares estimator for  $\mathbf{a}$  - from model (1) - is defined as

$$\hat{\mathbf{a}} = \operatorname{argmin}_{\mathbf{a}} \left\{ \sum_{i,g} (y_{i,g} - \mathbf{x}_g^\top \mathbf{a})^2 \right\} \quad (3)$$

which can also be written

$$\hat{\mathbf{a}} = \operatorname{argmin}_{\mathbf{a}} \left\{ \sum_{i,g} (y_{i,g} - \bar{y}_g + \bar{y}_g - \mathbf{x}_g^\top \mathbf{a})^2 \right\}. \quad (4)$$

Now, observe that

$$\begin{aligned} \sum_{i,g} (y_{i,g} - \bar{y}_g + \bar{y}_g - \mathbf{x}_g^\top \mathbf{a})^2 &= \sum_{i,g} (y_{i,g} - \bar{y}_g)^2 + (\bar{y}_g - \mathbf{X}_g^\top \mathbf{a})^2 \\ &\quad + 2(y_{i,g} - \bar{y}_g)(\bar{y}_g - \mathbf{x}_g^\top \mathbf{a}), \end{aligned}$$

where the first term is independent of  $\mathbf{a}$  (and can be removed from the optimization program), and the term with cross-elements sums to 0. Hence,

$$\hat{\mathbf{a}} = \operatorname{argmin}_{\mathbf{a}} \left\{ \sum_{i,g} (\bar{y}_g - \mathbf{x}_g^\top \mathbf{a})^2 \right\} = \operatorname{argmin}_{\mathbf{a}} \left\{ \sum_g n_g (\bar{y}_g - \mathbf{x}_g^\top \mathbf{a})^2 \right\} = \hat{\mathbf{b}}, \quad (5)$$

where  $\hat{\mathbf{b}}$  is the least square estimator of  $\mathbf{b}$  from model (2), when weights  $n_g$  are considered.

(ii) If we consider the sum of predicted values, observe that

$$\sum_{i,g} \hat{y}_{i,g} = \sum_g n_g \mathbf{x}_g^\top \hat{\mathbf{a}} = \sum_g n_g \underbrace{\mathbf{x}_g^\top \hat{\mathbf{b}}}_{\hat{\bar{y}}_g} = \sum_g \hat{y}_g. \quad (6)$$

Hence, the sum of predictions obtained from model (1) is the same as the sum of predictions obtained from model (2), even if partial sums are considered.

In the proposition above, the equality should be understood as the equality between estimators. Hence we have the following corollary.

**Corollary 2.2.** *We define the following matrices*

$$\begin{aligned}
\mathbf{Y}_g &= [Y_{1,g} \quad \dots \quad Y_{n_g,g}]^T & \mathbf{Y} &= [\mathbf{Y}_1^T \quad \dots \quad \mathbf{Y}_m^T]^T & \bar{\mathbf{Y}} &= [\bar{Y}_1 \quad \dots \quad \bar{Y}_m]^T \\
\boldsymbol{\varepsilon}_g &= [\varepsilon_{1,g} \quad \dots \quad \varepsilon_{n_g,g}]^T & \boldsymbol{\varepsilon} &= [\boldsymbol{\varepsilon}_1 \quad \dots \quad \boldsymbol{\varepsilon}_m]^T & \mathbf{e} &= [e_1 \quad \dots \quad e_m]^T \\
\mathbf{x}_g^{n_g} &= \underbrace{[\mathbf{x}_g \quad \dots \quad \mathbf{x}_g]}_{n_g \text{ times}} & \mathbf{x} &= [\mathbf{x}_1^{n_1} \quad \dots \quad \mathbf{x}_m^{n_m}] & \bar{\mathbf{x}} &= [\mathbf{x}_1^1 \quad \dots \quad \mathbf{x}_m^1],
\end{aligned}$$

the  $(1 \times n_g)$  vectors  $\mathbf{1}_{n_g} = [1 \quad \dots \quad 1]$  and  $\mathbf{0}_{n_g} = [0 \quad \dots \quad 0]$ , and the matrix

$$\mathbf{1} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_2} & \dots & \mathbf{0}_{n_m} \\ \mathbf{0}_{n_1} & \mathbf{1}_{n_2} & \dots & \mathbf{0}_{n_m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_1} & \mathbf{0}_{n_2} & \dots & \mathbf{1}_{n_m} \end{bmatrix}.$$

The OLS estimators are given by

$$\begin{aligned}
\hat{\mathbf{A}}_{OLS} &= \underset{\mathbf{a}}{\operatorname{argmin}} \{ (\mathbf{Y} - \mathbf{x}^T \mathbf{a})^T (\mathbf{Y} - \mathbf{x}^T \mathbf{a}) \} \\
&= (\mathbf{x} \mathbf{x}^T)^{-1} \mathbf{x} \mathbf{Y} \\
\hat{\mathbf{B}}_{OLS} &= \underset{\mathbf{b}}{\operatorname{argmin}} \{ ((\sqrt{\mathbf{1} \mathbf{1}^T}) \bar{\mathbf{Y}} - \sqrt{\mathbf{1} \mathbf{1}^T} \bar{\mathbf{x}}^T \mathbf{b})^T ((\sqrt{\mathbf{1} \mathbf{1}^T}) \bar{\mathbf{Y}} - (\sqrt{\mathbf{1} \mathbf{1}^T}) \bar{\mathbf{x}}^T \mathbf{b}) \} \\
&= (\bar{\mathbf{x}} \mathbf{1} \mathbf{1}^T \bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}} \mathbf{1} \mathbf{1}^T \bar{\mathbf{Y}}.
\end{aligned}$$

Model (1) on a micro level and model (2) on a macro level are equivalent, in the sense that

- (i)  $\mathbb{E}[\hat{\mathbf{A}}_{OLS}] = \mathbb{E}[\hat{\mathbf{B}}_{OLS}]$  and  $\operatorname{Var}[\hat{\mathbf{A}}_{OLS}] = \operatorname{Var}[\hat{\mathbf{B}}_{OLS}]$ , when weights  $n_g$  are used in model (2); and
- (ii)  $\mathbb{E} \left[ \sum_{i,g} \hat{Y}_{i,g} \right] = \mathbb{E} \left[ \sum_g \hat{Y}_g \right]$  and  $\operatorname{Var} \left[ \sum_{i,g} \hat{Y}_{i,g} \right] = \operatorname{Var} \left[ \sum_g \hat{Y}_g \right]$ .

*Proof.* Straightforward calculations lead to  $(\mathbf{1} \mathbf{1}^T)^{-1} \mathbf{1} \mathbf{Y} = \bar{\mathbf{Y}}$  and  $\bar{\mathbf{x}} \mathbf{1} = \mathbf{x}$ .

- (i) Let

$$\begin{aligned}
\mathbb{E} [\hat{\mathbf{B}}_{OLS}] &= (\bar{\mathbf{x}} \mathbf{1} \mathbf{1}^T \bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}} \mathbf{1} \mathbf{1}^T \mathbb{E} [\bar{\mathbf{Y}}] \\
&= (\bar{\mathbf{x}} \mathbf{1} \mathbf{1}^T \bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}} \mathbf{1} \mathbf{1}^T (\mathbf{1} \mathbf{1}^T)^{-1} \mathbf{1} \mathbb{E} [\mathbf{Y}] \\
&= (\mathbf{x} \mathbf{x}^T)^{-1} \mathbf{x} \mathbb{E} [\mathbf{Y}] = \mathbb{E} [\hat{\mathbf{A}}_{OLS}].
\end{aligned}$$

For the equality of variances, we have

$$\begin{aligned}
& \text{Var} \left[ \widehat{\mathbf{B}}_{OLS} \right] \\
&= (\bar{\mathbf{x}} \mathbf{1} \mathbf{1}^\top \bar{\mathbf{x}}^\top)^{-1} \bar{\mathbf{x}} \mathbf{1} \mathbf{1}^\top \text{Var} [\bar{\mathbf{Y}}] \left( (\bar{\mathbf{x}} \mathbf{1} \mathbf{1}^\top \bar{\mathbf{x}}^\top)^{-1} \bar{\mathbf{x}} \mathbf{1} \mathbf{1}^\top \right)^\top \\
&= (\bar{\mathbf{x}} \mathbf{1} \mathbf{1}^\top \bar{\mathbf{x}}^\top)^{-1} \bar{\mathbf{x}} \mathbf{1} \mathbf{1}^\top (\mathbf{1} \mathbf{1}^\top)^{-1} \mathbf{1} \text{Var} [\mathbf{Y}] \mathbf{1} \mathbf{1}^\top ((\mathbf{1} \mathbf{1}^\top)^{-1})^\top \left( (\bar{\mathbf{x}} \mathbf{1} \mathbf{1}^\top \bar{\mathbf{x}}^\top)^{-1} \bar{\mathbf{x}} \mathbf{1} \mathbf{1}^\top \right)^\top \\
&= (\mathbf{x} \mathbf{x}^\top)^{-1} \mathbf{x} \text{Var} [\mathbf{Y}] \mathbf{x}^\top ((\mathbf{x} \mathbf{x}^\top)^{-1})^\top \\
&= \text{Var} \left[ \widehat{\mathbf{A}}_{OLS} \right].
\end{aligned}$$

(ii) Let

$$\begin{aligned}
\mathbb{E} \left[ \sum_g \widehat{Y}_g \right] &= \mathbb{E} \left[ \mathbf{1}_m \mathbf{1} \mathbf{1}^\top \widehat{\bar{\mathbf{Y}}} \right] = \mathbb{E} \left[ \mathbf{1}_m \mathbf{1} \mathbf{1}^\top \bar{\mathbf{x}}^\top \widehat{\mathbf{B}} \right] \\
&= \mathbb{E} \left[ \mathbf{1}_m \mathbf{1} \mathbf{1}^\top \bar{\mathbf{x}}^\top \widehat{\mathbf{A}} \right] = \mathbb{E} \left[ \mathbf{1}_n \mathbf{x}^\top \widehat{\mathbf{A}} \right] \\
&= \mathbb{E} \left[ \mathbf{1}_n \widehat{\mathbf{Y}} \right] = \mathbb{E} \left[ \sum_{i,g} Y_{i,g} \right].
\end{aligned}$$

The proof of the equality of variances is similar.

**2.2. The quasi-Poisson regression.** A similar result can be obtained in the context of Poisson regressions. A generalized linear model [19] is made up of a linear predictor  $\mathbf{x}^\top \mathbf{b}$ , a link function that describes how the expected value depends on this linear predictor and a variance function that describes how the variance, depends on the expected value  $\text{Var}[Y] = \varphi V(\mathbb{E}[Y])$ , where  $\varphi$  denotes the dispersion parameter. For the Poisson model, the variance is equal to the mean, i.e.,  $\varphi = 1$  and  $V(\mathbb{E}[Y]) = \mathbb{E}[Y]$ . This may be too restrictive for many actuarial illustrations, which often show more variation than given by expected values. We use the term *over-dispersed* for a model where the variance exceeds the expected value. A common way to deal with over-dispersion is a *quasi-likelihood* approach (see [19] for further discussion) where a model is characterized by its first two moments.

Consider either a Poisson regression model, or a quasi-Poisson one,

$$Y_{i,g} \sim \mathcal{P}(\lambda_{i,g}) \text{ or } Y_{i,g} \sim q\mathcal{P}(\lambda_{i,g}). \quad (7)$$

In the case of a Poisson regression,

$$\mathbb{E}[Y_{i,g}] = \lambda_{i,g} = \exp[\mathbf{x}_g^\top \mathbf{a} + \ln(1/n_g)] \text{ and } \text{Var}[Y_{i,g}] = \lambda_{i,g}, \quad (8)$$

and in the context of a quasi-Poisson regression,

$$\mathbb{E}[Y_{i,g}] = \lambda_{i,g} = \exp[\mathbf{x}_g^\top \mathbf{a} + \ln(1/n_g)] \text{ and } \text{Var}[Y_{i,g}] = \varphi_{\text{micro}} \lambda_{i,g}, \quad (9)$$

with  $\varphi_{\text{micro}} > 0$  for a quasi-Poisson regression ( $\varphi_{\text{micro}} > 1$  for overdispersion). Here again, stacking observations within a cluster yield the following model (on the sum and not the average value, to have a valid interpretation with a Poisson distribution)

$$Y_g = \sum_i Y_{i,g} \sim \mathcal{P}(\lambda_g) \text{ or } Y_g \sim q\mathcal{P}(\lambda_g). \quad (10)$$

In the context of a Poisson regression,

$$\mathbb{E}[Y_g] = \lambda_g = \exp[\mathbf{x}_g^\top \mathbf{b}] \text{ and } \text{Var}[Y_g] = \lambda_g,$$

and in the context of a quasi-Poisson regression,

$$\mathbb{E}[Y_g] = \lambda_g = \exp[\mathbf{x}_g^\top \mathbf{b}] \text{ and } \text{Var}[Y_g] = \varphi_{\text{macro}} \lambda_g, \quad (11)$$

with  $\varphi_{\text{macro}} > 0$  for a quasi-Poisson regression. Here again, those two models ("micro" and "macro") are equivalent, in the sense that the following proposition holds.

**Proposition 2.3.** *Model (7) on a micro level and model (10) on a macro level are equivalent in the sense that*

- (i)  $\hat{\mathbf{a}}_{MLE} = \hat{\mathbf{b}}_{MLE}$ ; and
- (ii)  $\sum_{i,g} \hat{y}_{i,g} = \sum_g \hat{y}_g$ .

*Proof.* (i) Maximum likelihood estimator of  $\mathbf{a}$  is the solution of

$$\sum_{i,g} \left( \frac{y_{i,g} - \exp[\mathbf{x}_g^\top \mathbf{a}]}{\varphi_{\text{micro}}} \right) \mathbf{x}_g = \mathbf{0}$$

or equivalently

$$\sum_{i,g} (y_{i,g} - \exp[\mathbf{x}_g^\top \mathbf{a}]) \mathbf{x}_g = \mathbf{0}$$

With offsets  $\lambda_g^* = \exp[\mathbf{x}_g^\top \mathbf{b} + \log(n_g)]$ ,  $g = 1, \dots, m$ , maximum likelihood estimator of  $\mathbf{b}$  is the solution (as previously, we can remove  $\varphi_{\text{macro}}$ ) of



$$\begin{aligned}\sum_g (y_g - n_g \exp[\mathbf{x}_g^\top \mathbf{b}]) \mathbf{x}_g &= 0 \\ \sum_{i,g} (y_{i,g} - \exp[\mathbf{x}_g^\top \mathbf{b}]) \mathbf{x}_g &= 0.\end{aligned}$$

Hence,  $\hat{\mathbf{a}} = \hat{\mathbf{b}}$ , as (unique) solutions of the same system of equations.

(ii) The sum of predicted values is

$$\begin{aligned}\sum_{i,g} \hat{y}_{i,g} &= \sum_g n_g \hat{\lambda}_{i,g} = \sum_g n_g \exp[\mathbf{x}_g^\top \hat{\mathbf{a}}] = \sum_g n_g \exp[\mathbf{x}_g^\top \hat{\mathbf{b}}] \\ &= \sum_g \exp[\mathbf{x}_g^\top \hat{\mathbf{b}} + \log(n_g)] = \sum_g \hat{\lambda}_g^* = \sum_g \hat{y}_g.\end{aligned}$$

Nevertheless, as we will see later on, the Corollary obtained in the context of a Gaussian linear model does not hold in the context of a quasi-Poisson regression.

**Corollary 2.4.** *Model (7) on a micro level and model (10) on a macro level are asymptotically equivalent for Poisson regressions, in the sense that*

- (i)  $\mathbb{E}[\hat{\mathbf{A}}_{MLE}] = \mathbb{E}[\hat{\mathbf{B}}_{MLE}]$  and  $\text{Var}[\hat{\mathbf{A}}_{MLE}] = \text{Var}[\hat{\mathbf{B}}_{MLE}]$ , when  $n$  goes to infinity; and
- (ii)  $\mathbb{E}\left[\sum_{i,g} \hat{Y}_{i,g}\right] = \mathbb{E}\left[\sum_g \hat{Y}_g\right]$  and  $\text{Var}\left[\sum_{i,g} \hat{Y}_{i,g}\right] = \text{Var}\left[\sum_g \hat{Y}_g\right]$ , when  $n$  goes to infinity.

*Proof.* (i) A classical result of asymptotic theory for maximum likelihood estimators indicates that, under mild regularity conditions,  $\mathbb{E}[\hat{\mathbf{A}}_{MLE}] \rightarrow \mathbf{a}$  and  $\mathbb{E}[\hat{\mathbf{B}}_{MLE}] \rightarrow \mathbf{b}$  as  $n \rightarrow \infty$ . Without "micro" covariates in the model,  $\mathbf{a} = \mathbf{b}$  and so,  $\mathbb{E}[\hat{\mathbf{B}}_{MLE}] = \mathbb{E}[\hat{\mathbf{A}}_{MLE}]$  when  $n \rightarrow \infty$ . For model (7), the Fisher information matrix is  $\mathbf{I}(\mathbf{A}) = \mathbf{x}\mathbf{W}\mathbf{x}^\top$  and, when  $n \rightarrow \infty$ ,  $\text{Var}[\hat{\mathbf{A}}] \rightarrow (\mathbf{x}\mathbf{W}\mathbf{x}^\top)^{-1}$ , where  $\mathbf{W} = \text{diag}((\lambda_1/n_1)\mathbf{1}_{n_1}, \dots, (\lambda_m/n_m)\mathbf{1}_{n_m})$ . For model (10), we have  $\mathbf{I}(\mathbf{B}) = \bar{\mathbf{x}}\mathbf{1}\mathbf{W}\mathbf{1}^\top\bar{\mathbf{x}}^\top = \mathbf{x}\mathbf{W}\mathbf{x}^\top$  and, when  $n \rightarrow \infty$ ,  $\text{Var}[\hat{\mathbf{B}}] \rightarrow (\mathbf{x}\mathbf{W}\mathbf{x}^\top)^{-1}$ .

(ii) By using a similar argument, we have when  $n$  goes to infinity

$$\begin{aligned}\mathbb{E} \left[ \sum_g \hat{Y}_g \right] &= \mathbb{E} \left[ \mathbf{1}_m \mathbf{1} \mathbf{1}^\top \hat{\mathbf{Y}} \right] = \mathbf{1}_m \mathbf{1} \mathbf{1}^\top M_{\hat{\mathbf{B}}} (\bar{\mathbf{x}}^\top) \\ &= \mathbf{1}_m \mathbf{1} \mathbf{1}^\top M_{\hat{\mathbf{A}}} (\bar{\mathbf{x}}^\top) = \mathbb{E} \left[ \mathbf{1}_n \mathbf{1}^\top e^{\bar{\mathbf{x}}^\top \hat{\mathbf{A}}} \right] = \mathbb{E} \left[ \mathbf{1}_n e^{\mathbf{x}^\top \hat{\mathbf{A}}} \right] \\ &= \mathbb{E} \left[ \mathbf{1}_n \hat{\mathbf{Y}} \right] = \mathbb{E} \left[ \sum_{i,g} \hat{Y}_{g,i} \right].\end{aligned}$$

In small or moderate-sized samples, it should be noted that  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{B}}$  may be biased for  $\mathbf{A}$  and  $\mathbf{B}$ , respectively. Generally, this bias is negligible compared with the standard errors (see [20] and [21]).

In the quasi-Poisson micro-level model (from model (7)), as discussed above, the estimator of  $\mathbf{a}$  is the solution of the quasi-score function

$$\sum_{i,g} \left( \frac{y_{i,g} - \lambda_{i,g}}{\varphi_{micro}} \right) \mathbf{x}_g = 0,$$

which implies  $\hat{\mathbf{a}}_{QLE} = \hat{\mathbf{a}}_{MLE}$ . The classical Pearson estimator for the dispersion parameter  $\varphi_{micro}$  is

$$\hat{\varphi}_{micro} = \sum_{i,g} \frac{(y_{i,g} - \hat{y}_{i,g})^2 / \hat{y}_{i,g}}{\sum_g n_g - (k+1)} = \sum_{i,g} \frac{(n_g y_{i,g} - \hat{y}_g)^2 / n_g \hat{y}_g}{\sum_g n_g - (k+1)}.$$

Empirical evidence (see [29]) support the use of the Pearson estimator for estimating  $\varphi$  because it is the most robust against the distributional assumption. In a similar way, the quasi-Poisson macro-level model (from model (10)), the estimator of  $\mathbf{b}$  is the solution of

$$\sum_g \left( \frac{y_g - \lambda_g}{\varphi_{macro}} \right) \mathbf{x}_g = 0,$$

which implies here also  $\hat{\mathbf{b}}_{QLE} = \hat{\mathbf{b}}_{MLE}$ . The dispersion parameter  $\varphi$  is estimated by

$$\hat{\varphi}_{macro} = \sum_g \frac{(y_g - \hat{y}_g)^2 / \hat{y}_g}{m - (k+1)}.$$

Clearly,  $\hat{\varphi}_{micro} \neq \hat{\varphi}_{macro}$  involving the following results.

**Corollary 2.5.** *Model (7) on a micro level and model (10) on a macro level are not asymptotically equivalent for quasi-Poisson regressions, in the sense that*

- (i)  $\mathbb{E}[\hat{\mathbf{A}}_{QLE}] = \mathbb{E}[\hat{\mathbf{B}}_{QLE}]$  but  $\text{Var}[\hat{\mathbf{A}}_{QLE}] \neq \text{Var}[\hat{\mathbf{B}}_{QLE}]$ , when  $n$  goes to infinity; and
- (ii)  $\mathbb{E}\left[\sum_{i,g} \hat{Y}_{i,g}\right] = \mathbb{E}\left[\sum_g \hat{Y}_g\right]$  but  $\text{Var}\left[\sum_{i,g} \hat{Y}_{i,g}\right] \neq \text{Var}\left[\sum_g \hat{Y}_g\right]$ , when  $n$  goes to infinity.

*Proof.* (i) The property that variances are not equal is a direct consequence of classical results from the theory of generalized linear models (see [19]), since the covariance matrices of estimators are given by

$$\text{Var}[\hat{\mathbf{B}}] \rightarrow \hat{\varphi}_{macro} (\mathbf{x} \mathbf{W} \mathbf{x}^\top)^{-1}$$

and

$$\text{Var}[\hat{\mathbf{A}}] \rightarrow \hat{\varphi}_{micro} (\mathbf{x} \mathbf{W} \mathbf{x}^\top)^{-1}, \quad (12)$$

when  $n$  goes to infinity. Thus, covariance matrices of estimators are asymptotically equal for the Poisson regression model but differ for the quasi-Poisson model because  $\hat{\varphi}_{micro} \neq \hat{\varphi}_{macro}$ .

- (ii) Since the MLE and the QLE share the same asymptotic distribution (see [19]), the proof is similar to 2.4(ii).

**2.3. Poisson regression with random effect.** In the micro-level model described by Equation (7), observations made for the same event (subject) at different periods are supposed to be independent. Within-subject correlation can be included in the model by adding random, or subject-specific, effects in the linear predictor. In the Poisson regression model with random intercept, the between-subject variation is modeled by a random intercept  $\gamma$  which represents the combined effects of all omitted covariates.

Let  $Y_g^{(t)}$  represent the sum of all observations from subject  $t$ , in the cluster  $g$  and

$$\begin{aligned}\mathbb{E} [Y_g^{(t)} | \gamma_t] &= \exp[\mathbf{x}_g^\top \mathbf{a} + \gamma_t] \\ \text{Var} [Y_g^{(t)} | \gamma_t] &= \mathbb{E} [Y_g^{(t)} | \gamma_t] \\ \boldsymbol{\gamma} &= \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_T \end{bmatrix} \sim \mathbf{N}_T(\mathbf{0}, \sigma^2 \mathbf{I}),\end{aligned}$$

where  $\mathbf{I}$  is the  $(T \times T)$  identity matrix, and  $\mathbf{N}_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  the  $T$ -dimensional Gaussian distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . Straight-forward calculations lead to

$$\begin{aligned}\mathbb{E} [Y_g^{(t)}] &= \exp[\mathbf{x}_g^\top \mathbf{a} + \sigma^2/2] \\ \text{Var} [Y_g^{(t)}] &= \mathbb{E} [Y_g^{(t)}] (1 + \mathbb{E} [Y_g^{(t)}] (\exp[\sigma^2] - 1)).\end{aligned}$$

Hence,

$$\text{Var} [Y_g^{(t)}] > \mathbb{E} [Y_g^{(t)}], \quad \sigma^2 > 0. \quad (13)$$

This last equation shows that the Poisson regression model with random intercept leads to an over-dispersed marginal distribution for the variable  $Y_g^{(t)}$ . The maximum likelihood estimation for parameters requires Laplace approximation and numerical integration (see the Chapter 4 of [22] for more details). This model is a special case of multilevel Poisson regression model and estimation can be performed with various statistical softwares such as HLM, SAS, Stata and R (with package **lme4**).

One may be interested to verify the need of a source of between-subject variation. Statistically, it is equivalent to testing the variance of  $\gamma$  to be zero. In this particular case, the null hypothesis places  $\sigma^2$  on the boundary of the model parameter space which complicates the evaluation of the asymptotic distribution of the classical likelihood ratio test (LRT) statistic. From the very general result of [23], it can be demonstrated (see [25]) that the asymptotic null distribution of the LRT statistic is a 50/50 mixture of  $\chi_0^2$  and  $\chi_1^2$  as  $\sum_g n_g \rightarrow \infty$ . In this case, obtaining an equivalent macro-level model is of little practical interest since the construction of the variance-covariance matrix would require knowledge of the individual ("micro") data.

### 3. CLUSTERING AND LOSS RESERVING MODELS

A loss reserving macro-level model is constructed from data summarized in a table called run-off triangle. Aggregation is performed by

occurrence and development periods (typically years). For occurrence period  $i$ ,  $i = 1, 2, \dots, I$ , and for development period  $j$ ,  $j = 1, 2, \dots, I$ , let  $C_{i,j}$  and  $Y_{i,j}$  represent the total cumulative paid amount and the incremental paid amount, respectively with  $Y_{i,j} = C_{i,j} - C_{i,j-1}$ ,  $i = 1, \dots, I$ ,  $j = 2, \dots, I$ .

$$\begin{bmatrix} C_{1,1} & C_{1,2} & \dots & C_{1,I-1} & C_{1,I} \\ C_{2,1} & C_{2,2} & \dots & C_{2,I-1} & \\ \vdots & \vdots & \ddots & & \\ C_{I,1} & & & & \end{bmatrix}$$

$$\begin{bmatrix} Y_{1,1} = C_{1,1} & Y_{1,2} & \dots & Y_{1,I-1} & Y_{1,I} \\ Y_{2,1} = C_{2,1} & Y_{2,2} & \dots & Y_{2,I-1} & \\ \vdots & \vdots & \ddots & & \\ Y_{I,1} = C_{I,1} & & & & \end{bmatrix}$$

where columns, rows and diagonals represent development, occurrence and calendar periods, respectively. Each incremental cell  $Y_{i,j}$  can be seen as a cluster stacking  $n_{i,j}$  amounts paid in the same development period  $j$  for the occurrence period  $i$ . These payments come from  $M$  claims and let  $Y_{i,j}^{(k)}$  represent the sum of all observations from claims  $k$  in the cluster  $(i, j)$ . It should be noted that all claims are not necessarily represented in each of the clusters.

To calculate a *best estimate* for the reserve, the lower part of the triangle must be predicted and the total reserve amount is

$$\hat{R} = \sum_{t=2}^I \hat{C}_{t,I} - \sum_{t=2}^I C_{t,I-t+1} = \sum_{t=2}^I \sum_{s=I+2-t}^I \hat{Y}_{t,s}.$$

To quantify uncertainty in estimated claims reserve, we consider the mean square error of prediction (MSEP). Let  $\hat{R}$  be a  $\mathcal{Y}$ -mesurable estimator for  $\mathbb{E}[R|\mathcal{Y}]$  and a  $\mathcal{Y}$ -mesurable predictor for  $R$  where  $\mathcal{Y}$  represents the set of observed clusters. The MSEP is

$$\begin{aligned} MSEP_{R|\mathcal{Y}}(\hat{R}) &= \mathbb{E} \left[ \left( \hat{R} - R \right)^2 | \mathcal{Y} \right] \\ &= \text{Var} [R|\mathcal{Y}] + \left( \hat{R} - \mathbb{E} [R|\mathcal{Y}] \right)^2. \end{aligned}$$

Independence between  $R$  and  $\mathcal{Y}$  is assumed, so the equation is simplified as follows

$$MSEP_{R|\mathcal{Y}}(\widehat{R}) = \text{Var}[R] + \left(\widehat{R} - \mathbb{E}[R]\right)^2$$

and the unconditional MSEP is

$$MSEP_R(\widehat{R}) = \text{Var}[R] + \mathbb{E}\left[\left(\widehat{R} - \mathbb{E}[R]\right)^2\right].$$

### 3.1. The quasi-Poisson model for reserves.

3.1.1. *Construction.* From the theory presented in Subsection 2.2, we construct quasi-Poisson macro- and micro-level models for reserves. For both models, constitutive elements are defined in Table 1.

TABLE 1. Quasi-Poisson macro- and micro-level models for reserve  $(i, j = 1, \dots, I)$ . All clusters and all payments are independent.

Components	Macro	Micro
Exp. value	$\mathbb{E}[Y_{i,j}] = \lambda_{i,j}$	$\mathbb{E}[Y_{i,j}^{(k)}] = \lambda_{i,j}$
Inv. link func.	$\lambda_{i,j} = \exp[\mathbf{x}_{i,j}^\top \mathbf{b}]$ $= \exp[b_i + b_{I+j}]$ with $b_{I+1} = 0$	$\lambda_{i,j} = \exp[\mathbf{x}_{i,j}^\top \mathbf{a} + \log(1/n_{i,j})]$ $= \exp[a_i + a_{I+j} + \log(1/n_{i,j})]$ with $a_{I+1} = 0$
Variance	$\text{Var}[Y_{i,j}] = \varphi_{\text{macro}} \lambda_{i,j}$	$\text{Var}[Y_{i,j}^{(k)}] = \varphi_{\text{micro}} \lambda_{i,j}$
Pred. value	$\widehat{Y}_{i,j} = \exp[\widehat{b}_i + \widehat{b}_{I+j}]$	$\widehat{Y}_{i,j}^{(k)} = \exp[\widehat{a}_i + \widehat{a}_{I+j} + \log(1/n_{i,j})]$
Known values	$\mathcal{Y}_{\text{macro}}$	$\mathcal{Y}_{\text{micro}}$

As a direct consequence of Proposition 2.3, the best estimate for the total reserve amount is

$$\widehat{R} = \sum_{(i,j) \in \mathcal{K}} \sum_{k=1}^{n_{i,j}} \widehat{Y}_{i,j}^{(k)} = \sum_{(i,j) \in \mathcal{K}} \widehat{Y}_{i,j},$$

where  $\mathcal{K}$  represents unobserved clusters. For both models, the Proposition 3.1 gives results for the unconditional MSEP.

**Proposition 3.1.** *In the quasi-Poisson macro-level model, the unconditional MSEP is given by*

$$\begin{aligned} \widehat{MSEP}_R(\widehat{R}) &\approx \sum_{(i,j) \in \mathcal{K}} \widehat{\varphi}_{macro} \widehat{y}_{i,j} \\ &\quad + \sum_{(i,j), (n,m) \in \mathcal{K}} \widehat{\varphi}_{macro} \widehat{y}_{i,j} \widehat{y}_{m,n} \mathbf{x}_{i,j}^\top (\mathbf{x} \mathbf{W} \mathbf{x}^\top)^{-1} \mathbf{x}_{n,m}, \end{aligned}$$

where  $\mathbf{x}$  and  $\mathbf{W}$  are defined by Equation (12). The unconditional  $MSEP$  for the quasi-Poisson micro-level model is similar with  $\widehat{\varphi}_{macro}$  replaced by  $\widehat{\varphi}_{micro}$ .

*Proof.* The proof for the macro-level model is done in [21]. For the micro-level model, we have

$$\begin{aligned} MSE P_R(\widehat{R}) &= \text{Var}[R] + \mathbb{E} \left[ \left( \widehat{R} - \mathbb{E}[R] \right)^2 \right] \\ &= \sum_{(i,j) \in \mathcal{K}} \sum_{k=1}^{n_{i,j}} \widehat{\varphi}_{micro} \exp[\mathbf{x}_{i,j}^\top \widehat{\mathbf{a}} + \log(1/n_{i,j})] \\ &\quad + \sum_{(i,j) \in \mathcal{K}} \sum_{(m,n) \in \mathcal{K}} \sum_{k=1}^{n_{i,j}} \sum_{t=1}^{n_{m,n}} \text{Cov} \left[ \widehat{Y}_{i,j}^{(k)}, \widehat{Y}_{m,n}^{(t)} \right] \\ &= \sum_{(i,j) \in \mathcal{K}} \widehat{\varphi}_{micro} \exp[\mathbf{x}_{i,j}^\top \widehat{\mathbf{a}}] \\ &\quad + \sum_{(i,j) \in \mathcal{K}} \sum_{(m,n) \in \mathcal{K}} \sum_{k=1}^{n_{i,j}} \sum_{t=1}^{n_{m,n}} \exp[\mathbf{x}_{i,j}^\top \widehat{\mathbf{a}} + \log(1/n_{i,j})] \exp[\mathbf{x}_{m,n}^\top \widehat{\mathbf{a}} + \log(1/n_{m,n})] \\ &\quad \times \text{Cov} [\exp[\mathbf{x}_{i,j}^\top \widehat{\mathbf{a}} - \mathbf{x}_{i,j}^\top \mathbf{a}], \exp[\mathbf{x}_{m,n}^\top \widehat{\mathbf{a}} - \mathbf{x}_{m,n}^\top \mathbf{a}]] \end{aligned}$$

Although  $\widehat{Y}_{i,j}^{(k)}$  is not an unbiased estimator of  $\mathbb{E}[Y_{i,j}^{(k)}]$ , the bias is generally of small order and by using the approximation  $\exp[x] \approx 1 + x$  for  $x \approx 0$ , we obtain

$$\begin{aligned} &= \sum_{(i,j) \in \mathcal{K}} \widehat{\varphi}_{micro} \exp[\mathbf{x}_{i,j}^\top \widehat{\mathbf{a}}] \\ &\quad + \sum_{(i,j), (m,n) \in \mathcal{K}} \exp[\mathbf{x}_{i,j}^\top \widehat{\mathbf{a}} + \mathbf{x}_{m,n}^\top \widehat{\mathbf{a}}] \text{Cov} [\mathbf{x}_{i,j}^\top \widehat{\mathbf{a}}, \mathbf{x}_{m,n}^\top \widehat{\mathbf{a}}]. \end{aligned}$$

By using the fact that  $\widehat{\mathbf{b}} = \widehat{\mathbf{a}}$  and the remark at the end of subsection 2.2, we obtain

$$= \sum_{(i,j) \in \mathcal{K}} \widehat{\varphi}_{micro} \widehat{y}_{i,j} + \sum_{(i,j), (m,n) \in \mathcal{K}} \widehat{\varphi}_{micro} \widehat{y}_{i,j} \widehat{y}_{m,n} \mathbf{x}_{i,j}^\top (\mathbf{x} \mathbf{W} \mathbf{x}^\top)^{-1} \mathbf{x}_{m,n}.$$

Thus, the difference between the variability in macro- and micro-level models results from the difference between dispersion parameters. Define standardized residuals for both models

$$r_{i,g} = \frac{(y_{i,g} - \hat{y}_{i,g})}{\sqrt{\hat{y}_{i,g}}} \text{ and } r_g = \frac{(y_g - \hat{y}_g)}{\sqrt{\hat{y}_g}}.$$

Direct calculations lead to

$$\Psi = \frac{\sum_{i,g} r_{i,g}^2}{\sum_g r_g^2} \leq \frac{\sum_g n_g - (k+1)}{m - (k+1)} \rightarrow \hat{\varphi}_{\text{micro}} \leq \hat{\varphi}_{\text{macro}}. \quad (14)$$

Thus, if the total number of payments ( $\sum_g n_g$ ) is greater than the value  $\Psi(m - (k+1)) + k+1$ , then the micro-level model (7) will lead to a greater precision for the best estimate of the total reserve amount and conversely. Adding one or more covariate(s) at the micro level will decrease the numerator of  $\Psi$  and will increase the interest of the micro-level model.

**3.1.2. Illustration and Discussion.** To illustrate these results, we consider the incremental run-off triangle from UK Motor Non-Comprehensive account (published by [26]) presented in Table 2 where each cell  $(i, j)$ ,  $i + j \leq 7$ , is assumed to be a cluster  $g$ , i.e., the value  $Y_g$  is the sum of  $n_g$  independent payments.

TABLE 2. Incremental run-off triangle for macro-level model (in 000's).

	1	2	3	4	5	6	7
1	3 511	3 215	2 266	1 712	1 059	587	340
2	4 001	3 702	2 278	1 180	956	629	–
3	4 355	3 932	1 946	1 522	1 238	–	–
4	4 295	3 455	2 023	1 320	–	–	–
5	4 150	3 747	2 320	–	–	–	–
6	5 102	4 548	–	–	–	–	–
7	6 283	–	–	–	–	–	–

We construct 2 macro-level models

$$\textbf{Model A: } Y_g \sim \mathcal{P}(\lambda_g) \quad \lambda_g = \exp[\mathbf{x}_g^T \mathbf{a}]$$

$$\textbf{Model B: } Y_g \sim q\mathcal{P}(\lambda_g)$$



and 2 micro-level models

$$\begin{aligned} \textbf{Model C: } Y_{i,g} &\sim \mathcal{P}(\lambda_{i,g}) & \lambda_{i,g} &= \exp[\mathbf{x}_g^\top \mathbf{a} - \log(n_g)] \\ \textbf{Model D: } Y_{i,g} &\sim q\mathcal{P}(\lambda_{i,g}). \end{aligned}$$

The final reserve amount obtained from the Mack's model ([2]) is 28 655 773\$. To create micro-level datasets from the "macro" one, we perform the following procedure:

- (1) simulate the number of payments for each cluster assuming  $N_g \sim \mathcal{P}(\theta)$ ,  $g = 1, \dots, m$ ;
- (2) for each cluster, simulate a  $(n_g \times 1)$  vector of proportions assuming  $\boldsymbol{\omega}_g = [\omega_1 \ \dots \ \omega_{n_g}]^\top \sim \text{Dirichlet}(\mathbf{1})$ ,  $g = 1, \dots, m$ ;
- (3) for each cluster, define

$$\begin{bmatrix} Y_{1,g} \\ \vdots \\ Y_{n_g,g} \end{bmatrix} = \lfloor \boldsymbol{\omega}_g Y_g \rfloor, \quad g = 1, \dots, m;$$

- (4) adjust **Model C** and **Model D**; and
- (5) calculate the best estimate and the MSEP of the reserve.

For each value of  $\theta$ , we repeat this procedure 1 000 times and we calculate the average best estimate and the average MSEP. Those results are consistent with Corollaries 2.4, 2.5 and Proposition 3.1. For Poisson regression (**Model A** and **C**), results are similar. For micro-level models, convergence of  $\sqrt{MSEP}$  towards (11 622) is fast. For quasi-Poisson regression (**Model B** and **D**), Figure 1 shows  $\sqrt{MSEP}$  as a function of a expected total number of payments, for the portfolio. Above a certain level, (close to 3 400 here), accuracy of the "micro" approach exceed the "macro".

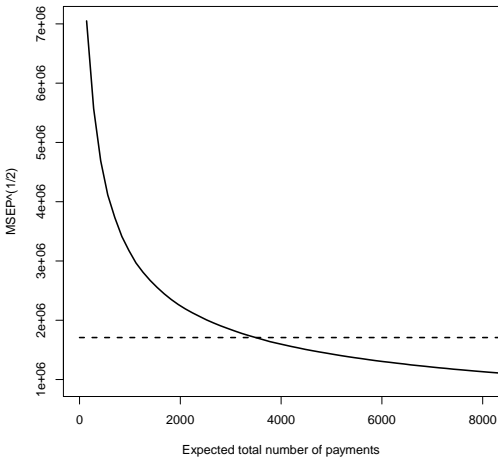


FIGURE 1. Square root of the mean square error of prediction obtained for **Model D** (solid line) and the **Model C** (broken line) from simulated values for increasing expected number of payments for the portfolio.

In order to illustrate the impact of adding a covariate at the micro-level, we define a quasi-Poisson micro-level model with a weakly correlated covariate (**Model E**) and with a strongly correlated covariate (**Model F**). Following a similar procedure, we obtain results presented in Table 3 and Figure 2.

TABLE 3. Results.		
Method	$\mathbb{E}[\text{Reserve}]$	$\sqrt{MSEP}$
Mack's model	28 655 773	1 417 267
Poisson reg.		
<b>Model A</b>	28 655 773	11 622
<b>Model C</b>	28 655 773	11 622
quasi-Poisson reg.		
<b>Model B</b>	28 655 773	1 708 196
<b>Model D</b>	28 655 773	see Figure 1
quasi-Poisson reg.		
<b>Model E</b> ( $\rho \approx 0$ )	28 657 364	see Figure 2
<b>Model F</b> ( $\rho \approx 0.8$ )	20 514 566	see Figure 2

As opposed to standard classical results on hierarchical models, the average of explanatory variable within a cluster  $((1/n_g) \sum_i x_{ig})$  has not been added to the macro-level model (**Model B**), for several reasons,

- (i) impossible to compute that average without individual data;
- (ii) discrete explanatory variables used in the micro-level model; and
- (iii) since claims reserve model have a predictive motivation, it is risky to project the value of an aggregated variable on future clusters.

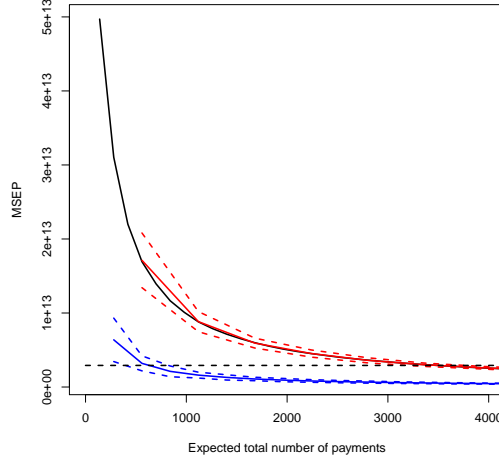


FIGURE 2. Mean square error of prediction ( $\pm 2\sigma$ ) obtained from simulated values as a function of the expected number of payments for **Model E** (red lines) and **Model F** (blue lines). For comparison purposes, the MSEP obtained for the **Model D** (solid black line) and the **Model B** (broken black line) are added.

With an explanatory variable highly correlated with the response variable, results obtained with **Model D** and **E** are very close. As claimed by Proposition 3.1 and equation (14), an explanatory variable highly correlated with the response variable will decrease the value of  $\sqrt{MSEP}$ , and lowers the threshold above which the micro-level model is more accurate than the macro-level one.

The quasi-Poisson macro-level model (**Model B**) with maximum likelihood estimators leads to the same reserves as the chain-ladder

algorithm and the Mack's model (see [28]), assuming the clusters exposure, for  $(i, j) \in \mathcal{K}$ , is one. To obtain similar results with a quasi-Poisson micro-level model (**Model D**), a similar assumption is necessary: exposure of each claim within cluster  $(i, j)$  is  $1/n_{i,j}$ . That assumption implies, on a micro level, that predicted individual payments  $\hat{Y}_{ij}^{(k)}$  are proportional to  $1/n_{i,j}$ . That assumption has unfortunately no foundation.

In the Poisson and quasi-Poisson micro-level models (**Model C** and **D**), payments related to the same claim, in two different clusters are supposed to be non-correlated. As discussed in the previous Section, it is possible to include dependencies among payments for a given claim using a Poisson regression with random effects. Simulations and computations were performed in R, using packages `ChainLadder` and `gtools`.

### 3.2. The Mixed Poisson model for reserves.

3.2.1. *Construction.* From the results obtained in Section 2.3, it is possible to construct a micro-model for the reserves that includes a random intercept. The later will allow to model dependence between payments from a given claim. Note that it is hard to find an aggregated model with random effects that could be compared with individual ones. In the context of claims reserves,  $Y_g^{(t)}$  represents the sum of paid made for claim  $t$  within cluster  $g$ . The assumptions of that model (called **model G**) are

$$\begin{aligned} (Y_g^{(t)} | \gamma_t) &\sim \mathcal{P}(\lambda_g e^{\gamma_t}), & \lambda_g &= \exp[\mathbf{x}_g^\top \mathbf{c} + \ln(1/n_g)] \\ \gamma_t &\sim N(0, \sigma^2). \end{aligned}$$

Because of those two random variables in the model, two kinds of predictions can be derived: un-conditional ones, where

$$\begin{aligned} (\hat{Y}_g^{(t)} | \gamma_t) &\sim \mathcal{P}(\hat{\lambda}_g e^{\gamma_t}) \\ \hat{\lambda}_g &= \exp[\mathbf{x}_g^\top \hat{\mathbf{c}} + \ln(1/n_g)] \\ \text{so that } \mathbb{E} [\hat{Y}_g^{(t)}] &= \lambda_g e^{\sigma^2/2}; \end{aligned}$$

and conditional ones, where the unknown magnitude of claim  $t$  is predicted by the so-called best linear estimate (that minimizes the MSEP)  $\tilde{\gamma}_t$  (see [21])

$$\begin{aligned} \left( \tilde{Y}_g^{(t)} | \tilde{\gamma}_t \right) &\sim \mathcal{P}(\hat{\lambda}_g e^{\tilde{\gamma}_t}) \\ \text{so that } \mathbb{E} \left[ \tilde{Y}_g^{(t)} \right] &= \lambda_g e^{\tilde{\gamma}_t}. \end{aligned}$$

It is then possible to compute the overall best estimate for the total amount of reserves.

**3.2.2. Illustration and Discussion.** In order to construct a micro-level model from triangle 2, we follow a procedure to the one described in the previous section, with steps 1–3 (that are not mentioned here)

4. for each accident year, allocate randomly the source of each payment
5. fit **model G**; and
6. compute the best estimate and the MSEP of the reserve.

For a fixed value of  $\theta$ , the procedure is repeated 1000 times. Various values were considered for  $\theta$  (10, 25, 50, 100 and 250), and results were similar. In order to avoid heavy tables, only the case where  $\theta = 10$  is mentioned here. Simulations and computations were performed with R, relying on package **lme4**. On Figure 3 we can see predictions of the model on observed data, while on Figure 4 we can see predictions of the model for non-observed cells. Finally, results are reported in Table 4. At each step, a LRT is performed (see section 2.3) and each time, the variance at origin was significant non-null, meaning that correlation among payments (related to the same claim) is positive. Observe that with the random model, the log-likelihood is approximated using numerical intergration, which might bias computed  $p$ -values of the test. Here,  $p$  have been confirmed using a bootstrap procedure (using package **glmmML**).

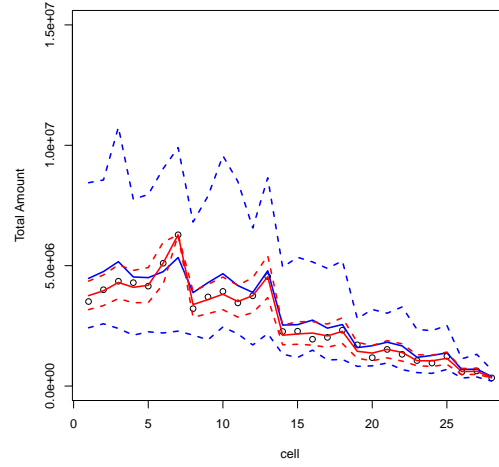


FIGURE 3. Observed data (circles) with conditional prediction (plain red lines) and un-conditional (plain blue lines) from **model G** ( $\theta = 10$ ).

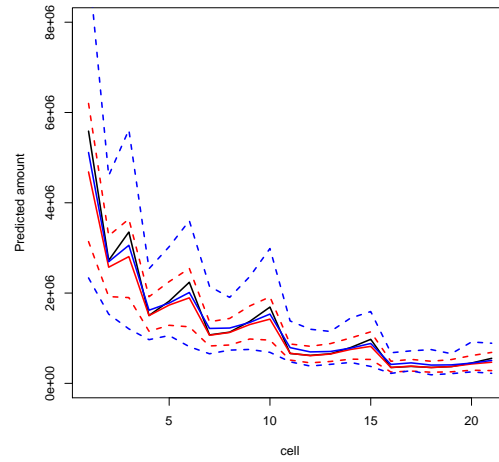


FIGURE 4. Predictions with the quasi-Poisson macro-level model (strong black line), with conditional prediction (plain red lines) and un-conditional (plain blue lines) from **model G** ( $\theta = 10$ ).

TABLE 4. Numerical Results for  $\theta = 10$ . Results for different values of  $\theta$  are similar.

Modèle	$\mathbb{E}[\text{Reserve}]$	$\sqrt{\text{Var}(\text{Reserve})}$
coll. quasi-Pois.	28 656 423	1 708 216
mixed Poisson non-cond.	27 930 624	3 297 401
mixed Poisson cond.	25 972 947	2 280 902

## REFERENCES

- [1] Astesan, E. *Les réserves techniques des sociétés d'assurances contre les accidents automobiles*; Librairie générale de droit et de jurisprudence, 1938.
- [2] Mack, R. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin* **1993**, 23(2), 213-225.
- [3] England, P.D.; Verrall, R.J. Stochastic claims reserving in general insurance. *British Actuarial Journal* **2003**, 8(3), 443-518.
- [4] van Eeghen, J. Loss reserving methods. In *Surveys of Actuarial Studies 1*; Nationale-Nederlanden, 1981.
- [5] Taylor, G.C. *Claims reserving in non-life insurance*; North-Holland, 1989.
- [6] Karlsson, J.E. (1976). Karlsson, J.E. The expected value of IBNR claims. *Scandinavian Actuarial Journal* **1976**, 1976(2), 108-110.
- [7] Arjas, E. The claims reserving problem in nonlife insurance - some structural ideas. *ASTIN Bulletin* **1989**, 19(2), 139-152.
- [8] Jewell, W.S. (1989). Jewell, W.S. Predicting IBNYR events and delays I. Continuous time. *ASTIN Bulletin* **1989**, 19(1), 25-55.
- [9] Norberg, R. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin* **1993**, 23(1), 95-115.
- [10] Hesselager, O. A Markov model for loss reserving. *ASTIN Bulletin* **1994**, 24(2), 183-193.
- [11] Norberg, R. Prediction of outstanding liabilities II: model variations and extensions. *ASTIN Bulletin*, **1999**, 29(1), 5-25.
- [12] Hesselager, O.; Verrall, R.J. *Reserving in Non-Life Insurance*; Wiley StatsRef: Statistics Reference Online, 2014.
- [13] Friedland, J. *Estimating Unpaid Claims Using Basic Techniques*; Estimating Unpaid Claims Using Basic Techniques, 2010.
- [14] Antonio K.; Plat R. Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal* **2014**, 2014(7), 649-669.
- [15] Pigeon M.; Antonio K.; Denuit M. Individual Loss Reserving using Paid-Incurred Data. *Insurance: Mathematics and Economics* **2014**, 58, 121-131.
- [16] van den Berga, G.J.; van der Klaauw, B. Combining micro and macro unemployment duration data. *Journal of Econometrics* **2001**, 102, 271-309.
- [17] Altissimo, F.; Mojon, B.; Zaffaroni, P. Fast micro and slow macro: can aggregation explain the persistence of inflation? *European Central Bank Working Papers* **2007**, 729.
- [18] Greene, W.H. *Econometric Analysis, Fifth edition*; Prentice Hall, 2003.
- [19] McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; Chapman

- [20] Cordeiro, G.M.; McCullagh, P. Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* **1991**, *53*(3), 629-643.
- [21] Wüthrich, M.; Merz, M. *Stochastic claims reserving methods*; Wiley Interscience, 2008.
- [22] Snijders, T.A.B.; Bosker, R.J. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*; Sage Publishing, 2012.
- [23] Self, S.G.; Liang, K.Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **1987**, *82*, 605-610.
- [24] Skrondal, A.; Rabe-Hesketh, S. Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society A* **2009**, *172*(3), 659-687.
- [25] Zhang, D.; Lin, X. Random Effect and Latent Variable Model Selection. *Lecture notes in statistics*, **2008**, 182.
- [26] Christofides, S. Regression models based on log-incremental payments. *Claims Reserving Manual* **1997**, D5.1-D5.53.
- [27] Imbens, G.W.; Lancaster, T. Combining micro and macro data in microeconomic models. *Review of Economic Studies* **1994**, *61*, 655-680.
- [28] Mack, T.; Venter, G. A Comparison of Stochastic Models that Reproduce Chain Ladder Reserve Estimates. *Insurance: Mathematics and Economics* **2000**, *26*, 101-107. & Hall/CRC, 1989.
- [29] Ruoyan, M. Estimation of dispersion parameters in GLMs with and without random effects. *Stockholm University* **2004**, <http://www2.math.su.se/matstat/reports/serieb/2004/rep5/report.pdf>.